Brain Topogr (2013) 26:110–125
DOI 10.1007/s10548-012-0237-y

ORIGINAL PAPER

Musicianship Boosts Perceptual Learning of Pseudoword-Chimeras: An Electrophysiological Approach

Jürg Kühnis · Stefan Elmer · Martin Meyer · Lutz Jäncke

Received: 20 December 2011 / Accepted: 7 June 2012 / Published online: 27 June 2012

Abstract A vast amount of previous work has consistently revealed that professional music training is associated with functional and structural alterations of auditory-related brain regions. Meanwhile, there is also an increasing array of evidence, which shows that musicianship facilitates segmental, as well as supra-segmental aspects of speech processing. Based on this evidence, we addressed a novel research question, namely whether professional music training has an influence on the perceptual learning of speech sounds. In the context of an EEG experiment, we presented auditory pseudoword-chimeras, manipulated in terms of spectral- or envelope-related acoustic information, to a group of professional musicians and non-musicians. During EEG measurements, participants were requested to assign the auditory-presented pseudoword-chimeras to one out of four visually presented templates. As expected, both groups showed behavioural learning effects during the time course of the experiment. These learning effects were associated with an increase in accuracy, a decrease in reaction time, as well as a decrease in the P2-like microstate duration in both groups. Notably, the musicians showed an increased learning performance compared to the controls during the first two runs of the spectral condition. This perceptual learning effect, which varies as a function of musical expertise, was reflected by a reduction of the P2-like microstate duration. Results may mirror transfer effects from musical training to the processing of spectral information in speech sounds. Hence, this study provides first evidence for a relationship between changes in microstates, musical expertise, and perceptual verbal learning mechanisms.

Keywords EEG · Microstates · Topographical pattern analysis · Auditory chimeras · Perceptual learning · Musical expertise · Plasticity

J. Kühnis (✉) · S. Elmer · L. Jäncke
Division of Neuropsychology, Institute of Psychology, University of Zurich, Binzmühlestrasse 14/25, 8050, Zurich, Switzerland
e-mail: j.kuehnis@psychologie.uzh.ch

S. Elmer
e-mail: s.elmer@psychologie.uzh.ch

M. Meyer · L. Jäncke
Center for Integrative Human Physiology, Zurich, Switzerland
e-mail: martin.meyer@uzh.ch

M. Meyer
Research Unit for Plasticity and Learning in the Healthy Aging Brain (HAB LAB), University of Zurich, Zurich, Switzerland

L. Jäncke
International Normal Aging and Plasticity Imaging Center (INAPIC), Zurich, Switzerland
e-mail: l.jaencke@psychologie.uzh.ch

Introduction

In the last two decades, neuroscientific evidence has demonstrated that the mammalian auditory system is highly susceptible to functional and structural changes (Jancke 2009; Munte et al. 2002; Schlaug 2001). For example, several animal studies provide evidence for cortical reorganization in primary auditory areas of adult cats after cochlear lesions (Rajan et al. 1993) or in the monkeys' auditory cortex after a frequency discrimination training (Recanzone et al. 1993). In the field of human neuroscience, the influence of intense musical training on auditory processing has been well documented in various cross-sectional (Baumann et al. 2008; Pantev et al. 2001a;

Meyer et al. 2011; Schlaug et al. 1995) and longitudinal studies (Moreno et al. 2009; Hyde et al. 2009). Since professional musicians usually start practising a musical instrument at an early stage of their life and usually spend many hours per day playing their musical instrument, this specific group of individuals is particularly fruitful for investigating neuroplastic alterations of the auditory system (Munte et al. 2002; Jancke 2009).

Previous evidence shows that musical training not only facilitates the processing of musical information, but that it also effects perceptual processing in domains that have not been explicitly trained (Besson et al. 2011; Patel 2011; Milovanov and Tervaniemi 2011; Schön and Francois 2011; Kraus and Chandrasekaran 2010). In fact, even though speech and music are characterized by different psychoacoustic features, both signals convey information by means of timing, pitch, and timbre cues (Kraus and Chandrasekaran 2010). Since there are more acoustic similarities between these two domains than differences, a growing number of neuroscientists are interested in investigating transfer effects from musical training to the processing of speech sounds. Most of the studies investigating such transfer effects address the question whether musical expertise may favour segmental (Marie et al. 2011; Ott et al. 2011; Elmer et al. 2012) or supra-segmental (Oechslin et al. 2010; Marie et al. 2011; Schon et al. 2004; Marques et al. 2007) aspects of elemental speech processing.

To date, there is also growing evidence indicating that musicianship not only favours perceptual processes, but that it is indeed associated with behavioural advantages in several cognitive domains (Baumann et al. 2008). Such benefits are apparent even in individuals who have not been explicitly trained (Aleman et al. 2000; Bilhartz et al. 1999; Brochard et al. 2004; Sluming et al. 2007; Ho et al. 2003). A research question that is of fundamental interest in association with musical training and transfer effects is whether musicianship exerts an influence on perceptual learning mechanisms. Even though a comparable research question has been previously addressed by other authors (Ho et al. 2003; Schlaug et al. 2005), it remains unclear exactly which of the trained acoustic parameters (i.e., frequency, amplitude, timing, or timbre) are relevant for facilitating transfer effects from musical training to the perceptual learning of speech sounds. Based on previous work illustrating a more efficient processing of spectral information in musicians (Besson et al. 2007; Boh et al. 2011; Draganova et al. 2009; Gaab et al. 2005; Schon et al. 2004), one would expect that musicians will have a behavioural advantage over non-musicians when it comes to learn auditory-presented verbal material that is characterized by spectral information only. To date, it is unknown whether musicians show behavioural advantages when speech-stimuli are characterized by envelope-related acoustic information only.

Even though the electrophysiological (EEG) technique has a long tradition in the field of cognitive neuroscience, recent novel applications and mathematical implementations have enabled researchers to address more complex questions in order to uncover how the human brain functions. Regarding the auditory system, most of the EEG studies focussed on the auditory-related N1–P2 complex or the associated latency bands which are known to reflect the auditory representation and energy of speech and non-speech sounds (Ostroff et al. 1998; Pratt et al. 2007; Sharma et al. 2000; Meyer et al. 2006). For example, previous EEG studies demonstrate an increased auditory representation of musical sounds in musicians compared to non-musicians (Meyer et al. 2011; Baumann et al. 2008; Besson et al. 2007; Schlaug et al. 2005). Results pointing to increased cortical representation of the primary instrument played by musicians have also come to light (Pantev et al. 2001b; Schneider et al. 2005a; Meyer et al. 2011).

A particularly interesting approach known as topographic pattern analysis was recently proposed by Murray and colleagues (2008). This methodological procedure is helpful in that it permits to evaluate and interpret EEG data by performing a step-by-step analysis. Even though the general principles of the microstate segmentation have been described extensively elsewhere (for a comprehensive overview consider Murray et al. 2008; Michel et al. 2009), in turn we will provide a short summary of this procedure. The topographic pattern analysis can be used for evaluating complex topographic information arising from high-density EEG measurements. In particular, it permits to determine electrophysiological differences between experimental conditions and groups with regard to field strength, topography, and latency shift. This specific method implies that evoked activities are more or less stable scalp voltage topography manifestations lasting for several tens of milliseconds before a qualitative change in the configuration of the electric fields occurs. Such a qualitative change leads to another topographic map, which remains stable for a certain duration. These stable topographic maps are called microstates and can be determined by using a topographic atomize and agglomerate hierarchical cluster analysis (T-AAHC) on the grand-averaged ERP's across conditions and groups.

In the present work we performed an EEG study specifically designed to evaluate perceptual learning mechanisms as a function of musical expertise. With this purpose in mind we presented pseudoword-chimeras characterized by spectral- and envelope-reduced information to musically trained and untrained individuals. In accordance with previous studies showing that musicians are principally facilitated in processing spectral information (Besson et al.

2007; Boh et al. 2011; Draganova et al. 2009; Gaab et al. 2005; Schon et al. 2004), we expected to obtain the following results: musicians will generally show an increased learning performance while recognizing the auditory chimeras that are only defined by spectral information. In addition, we expect that musicians will exhibit a faster learning curve during the time course of the experiment. Due to our consideration of previous work indicating that musicianship favours the processing of several different psychoacoustic dimensions (Kraus and Chandrasekaran 2010), such as, frequency- (Meyer et al. 2011), duration-, as well as amplitude-modulations (Lee et al. 2009), we also expect to find that musicians are likewise better at recognizing pseudoword-chimeras that are only characterized by amplitude information only.

With regard to the electrophysiological data, we expect to find increased microstate-related global field power (GFP) magnitudes of the N1–P2 complex in musicians in response to both spectral- and envelope-reduced pseudoword-chimeras. Furthermore, we hypothesize that the better learning performance of the musicians shall be reflected by an alteration of the N1–P2-like microstates duration. The reason for focussing on the N1–P2 complex is supported by the fact that these auditory-evoked responses have been shown to constitute an objective method for measuring the processing of any discrete acoustic features (Hyde 1997). In this context, previous work has consistently shown that musicianship is associated with increased cortical representations of speech (Marie et al. 2011) and non-speech sounds (Pantev et al. 1998), as reflected by enhanced brain responses. Consequently, one may expect that changes in any discrete acoustic features of the stimuli, as well as expertise, will be associated with a modulation of the microstates duration during the processing stage of the N1–P2 complex. This expectation is supported by earlier work indicating a relationship between microstates duration and brain processes (Lehmann and Michel 2011).

Materials and Methods

Participants

Fifty volunteers with normal audiological status and no history of neurological diseases participated in this study. All participants were native Swiss–German or German speakers and were consistently right-handed according to standard handedness questionnaires (Annett 1970; Jancke 1996). The first group of participants consisted of 25 professional musicians (13 men and 12 women) in the age range of 19 to 47 years (mean age 25.3 ± 5.9 SD). Musicians had more than 10 years of musical training. All musicians commenced their musical training before the age of 7 years (mean age 6.5 year ± 1.3 SD) and practised their musical instrument for 3.2 h/day on average (± 1.5 SD). The musician group consisted of seven singers, five pianists, four participants who play classical drums, and nine participants who play other instruments. The control group consisted of 25 volunteers without formal musical education in the age range from 19 to 45 years (mean age 26.4 ± 7.5 SD). All participants gave informed written consent in accordance with procedures approved by the local ethics committee and were paid for their participation.

Behavioural Data

Musical Aptitudes

In order to measure the musical aptitudes of each participant, we applied the "Advanced Measure of Music Audition" (AMMA) test implemented by Gordon (1989). This specific test is based on the assumption that holding auditory-presented musical patterns in memory and detecting both melodic and rhythmic variations constitute a fundamental prerequisite for musical aptitude. During the AMMA test volunteers heard short pairs of piano sequences (both pieces have a duration of about 2 s) and were required to decide whether these sequences were equivalent, rhythmically different, or tonally different.

Cognitive Capability

In order to exclude that putative differences in intelligence between the two groups could influence the data in some direction, we adopted two short intelligence tests: the KAI (Lehrl et al. 1991) and the MWT-B (Lehrl 1977). Table 1 gives an overview of the biographical and behavioural data of the participants.

Stimuli

The auditory stimuli we presented to the participants consisted of forty phonotactically legal low associative pseudowords taken from the verbal/non-verbal (VLT/NVLT) learning test (Sturm and Willmes 1999). All stimuli used in this test are complex disyllabic pseudowords (example: "hentes", "fasenz", "trelas"). The stimuli were spoken by a professional male speaker and recorded at a sampling rate of 44.1 kHz. All pseudowords were matched in intensity (amplitude normalization with the PRAAT software, http://www.fon.hum.uva.nl/praat/); the duration of the auditory stimuli ranged from 900 to 1,000 ms.

Table 1 Biographical and behavioural data of the two groups

	Age (a)		Gender		KAI (IQ)		MWT-B (IQ)		AMMA** (PR)	
	M	SD	*M*	*F*	*M*	SD	*M*	SD	*M*	SD
M	25.28	5.669	12	13	127.76	10.82	113.10	13.58	69.16	14.72
NM	26.44	7.556	12	13	123.74	12.22	115.48	14.21	55.76	16.82

M Musician, *NM* non-musicians, *AMMA* advanced measures of music aptitude, *KAI* Kurztest der aktuellen geistigen Leistungsfähigkeit, *MWT* Mehrfachwahl-Wortschatz-Intelligenz). ** Significant difference between the two groups, $P = 0.004$ ($F_{(1,\ 48)} = 8.984$)

For each of the auditory stimuli, we produced four different speech-noise chimeras by using the MATLAB software (versionR2010a http://www.mathworks.com). The first step was to create noise analogues of the original stimuli by maintaining the spectral density and intensity. The second step consisted of filtering both sound signals (i.e., the original stimuli and the noise analogues we created) with two, three, or eight complementary band-pass filters in the frequency range from 80 Hz to 17.64 kHz (0.4 × sampling frequency). This procedure was done by using nearly rectangular frequency filters with logarithmic frequency spacing (Shen et al. 2001; Smith et al. 2002). The purpose of this procedure was to diversify the intelligibility of the auditory stimuli by manipulating the number of band-pass filters (Smith et al. 2002). In the third processing step, the filtered signals from the corresponding frequency bands of the two signals (i.e., speech- and noise-stimuli) were converted by using a chimerizer procedure that interchanges the envelope (i.e., the time course of the signal amplitude) and the fine structure (i.e., the time course of the spectral information) of the two input signals by producing two partial speech-noise chimeras. We used the Hilbert Transform to factorize envelope and fine structure from the two input signals (speech- and noise-stimuli). With this purpose in mind, we modulated the fine structure and the envelope of the speech and noise signals (and vice versa), in order to obtain partial speech-noise chimeras (two partials chimeras per frequency band in total). Finally, we summed up these partials chimeras over all frequency bands, resulting in two multi-band chimeras for each of the three band-pass filtered input signals (i.e., two, three, or eight band-pass filters). Whereas the first chimera was characterized by the envelope of the speech sound and by the fine structure of the noise signal, the second one was composed of the envelope of the noise sound, a straight line, and the fine structure of the speech signal (Shen et al. 2001). A sample of these auditory stimuli can be downloaded under: (http://www.neurowissenschaft.ch/mmeyer/.

The visually presented words (one target and three distractors) consisted of pseudowords, which were phonological neighbours of the target word. These phonological neighbours were collected in the context of a pre-test study conducted with four students who were instructed to write down the pseudowords they heard. In this context, we used the erroneously reported pseudowords as distractors for the main experiment. By applying this procedure, we only focused on distractors that exhibited a similar phonological structure (example: target word: “hentes”; distractors: “hintis”, “temtis”, “sentis”).

Before running the EEG experiment, we performed a short behavioural experiment with 17 volunteers, in order to estimate the intelligibility of the speech-noise chimeras. Whereas all of the auditory pseudoword chimeras were presented via headphones, the target words and the distractors were visually presented on a computer screen. In this context, volunteers were instructed to recognize the pseudoword chimera that was presented auditorily and to select the visual target stimulus by pressing one of four response buttons. This procedure was the same as the method we employed in the EEG experiment. The only difference was that we recruited two different groups of individuals for performing the pilot study and the EEG experiment.

This behavioural approach revealed an intelligibility (i.e., number of correctly recognized items) of about 37 % for the chimeras containing speech information in the envelope and filtered with three band filters (ENV3), as well as for the chimera containing speech information in the fine structure and filtered with three band filters (FS3). The other two chimeras, namely, the chimera that was defined by the speech information in the envelope and filtered with two band filters (ENV2), as well as the chimera containing speech information in the fine structure and filtered with eight band filters (FS8) exhibited an intelligibility of about 30 %. As expected, the original and not the manipulated pseudowords were characterized by an intelligibility of nearly 100 %.

Experimental Procedure and Task

During the EEG measurement, participants sat in a comfortable chair at a distance of about 75 cm from a 19-inch-monitor and were instructed to look at a small fixation cross that was presented on the screen. In order to avoid

movement-related artefacts during data acquisition. Every participant performed four runs, each of which lasted about 15 min. During each run, the stimuli of all five conditions (ENV2, FS8, ENV3, FS3, and original pseudowords that acted as a control condition) were presented 160 times in a randomized order at a sound pressure level of about 75 dB (Digital Sound Level Meter 329, Voltcraft) with Sennheiser in-ear HIFI-headphones (CX-350). Two seconds after stimulus onset, the pseudowords (target word) and three other pseudowords, which acted as distractors were visually presented on the screen. Participants were instructed to recognize and select the pseudoword (target word) they heard by pressing one of four possible response buttons with the middle- and forefingers of both hands (randomized across the participants).

After response selection, a delay of one second followed before the next auditory stimulus was presented. The presentation of the auditory stimuli and the recording of behavioural responses were controlled by the Presentation software (www.neurobs.com; version 14.5). We collapsed the behavioural and electrophysiological responses of the FS3 and FS8 into one FS condition for all of the analysis in this work. This was done for two reasons: we were interested in analysing the influence of expertise on the processing of spectral- and enveloped-reduced pseudowords, and in order to increase the variance across the two groups. The same procedure was applied for the ENV3 and ENV2 stimuli (ENV condition).

EEG Acquisition and Pre-Processing

Continuous EEG was recorded by using a high-density Geodesic EEG system (Electrical Geodesics, Inc., USA) with 256 scalp electrodes (impedance <40 kΩ, online vertex reference; sampling rate of 500 Hz; band-pass filtered at 0.1–250 Hz). Brain Vision Analyser software (Version 2.01, Brainproducts, Munich, Germany) was used for all pre-processing steps. We reduced the number of electrodes from 256 to 204 by removing the electrodes situated on the cheeks and on the neck. Data were filtered off-line from 0.1 to 150 Hz. Artefacts (i.e., eye movements and eye blinking) were eliminated by using an independent component analysis (ICA) (Jung et al. 2000) in association with a semi-automatic raw data inspection. Furthermore, noisy data collected from distinct electrodes were interpolated (Perrin et al. 1987) and data were re-referenced off-line to an average reference, and sectioned into 500 ms segments (from 100 ms pre-stimulus to 400 ms post-stimulus). A baseline correction relative to the -100 to 0 ms pre-stimulus time period was applied and all segments were averaged for each participant, each condition, and each run, in order to compute auditory evoked potentials (AEPs). In addition, multi-subject grand averages were calculated for each group and condition.

Data Analysis

Behavioural Data

In a first step, we checked whether participants correctly performed the recognition task by averaging the responses to the clean, presented pseudowords for each participant. All participants attained recognition performances between 88 and 100 % (mean 96.5 %).

Following the lead of previous studies (Luo and Poeppel 2007; Shannon et al. 1995; Smith et al. 2002), we collapsed the number of correct ENV2 and ENV3 responses (accuracy ENV condition), as well as of the FS8 and FS3 responses (accuracy FS condition) for each participant and run. This was done in order to estimate general group differences while processing "envelope" (ENV) and "fine structure" (FS) stimuli. The same procedure was applied for evaluating reaction time (RT).

The evaluation of RT was performed by using a mixed-effect regression model analysis. Otherwise, since accuracy data were binomially distributed, we performed a generalized linear mixed model for binomially distributed outcome. For both analyses, we used *run* (run 1, 2, 3, 4), *condition* (ENV, FS), and *group* (musician, non-musician) as fixed factors, as well as *participants* and *items* as random factors (Baayen et al. 2008; Locker et al. 2007; Clark 1973). The main problem when using standard ANOVAs and regression models is that these analyses are conducted either by-subjects or by-items. By contrast, the mixed-effects model allows to include subject and item effects in the same model. This in turn potentially solves the language-as-fixed-effect fallacy (See Brysbaert (2007) for a primer on mixed-effect models). The model was then fitted using the restricted maximum likelihood estimation (REML).

EEG Data

In the present work ERP data were analysed by using a multistep procedure which bases on the analysis of response topography and response strength. This procedure is useful to differentiate between underlying generators as a function of expertise and task. Furthermore, this approach enables to disentangle latency shifts of brain activity across experimental conditions and groups from modulation in the strength of responses of statistically indistinguishable brain generators (Michel et al. 2009; Murray et al. 2008). This kind of analysis is independent of the reference electrode

(see, e.g. Michel et al. 2009) and is insensitive to amplitude modulations. All EEG analyses were performed by using the Cartool software package (Version 3.43/869; The Cartool Community group, sites.google.com/site/cartool community).

Global Field Power Analyses (GFP)

In a first evaluation step, which involves defining the time windows of interest, we compared GFP differences between the two groups by collapsing both conditions (ENV, FS) for each participant together into one auditory-evoked-potential (AEP) and by computing GFP time courses. In order to control for inter-individual variance across participants, we normalized the data to the mean GFP for each participant. For each electrode we marked the first time point when the previously calculated *t* test between the two groups exceeded the statistical alpha criterion of 0.05 for at least ten consecutive data points (i.e., time windows which were larger than 20 ms). This criterion was previously proposed to be an adequate measure for describing AEP modulations (Murray et al. 2002; Murray et al. 2004; Fort et al. 2002).

Microstate Analyses

In a second step, we performed microstate analyses. The logic behind this analysis was to identify spatio-temporal differences in scalp voltage distributions between the two groups and the two conditions (i.e., FS and ENV). With this purpose in mind, we applied a "topographic atomize and agglomerate hierarchical cluster" analysis (T-AAHC) on the grand-averaged AEP's for each condition and group over a segment length of 400 ms. The T-AAHC analysis is used for the temporal segmentation of the EEG data. It permits to identify periods of stable electric field topographies or "functional microstates". This segmentation procedure generates hypotheses about when topographic modulations actually occur. The periods of stable topography are functionally relevant periods of interest (https://sites.google.com/site/cartoolcommunity/). This procedure incorporates Krzanowski–Lai criterion to identify the optimal number of template maps that can explain the data (Murray et al. 2008; Tibshirani and Walther 2005; Krzanowski and Lai 1988).

The time window we used for statistical analysis primarily bases on the hypothesis that the processing of the perceptual properties of acoustic stimuli as a function of expertise is most likely reflected by a modulation of the N1–P2 complex. In order to validate this working hypothesis we additionally performed GFP comparisons between the two groups. Since the GFP analysis confirmed a modulation of brain responses overlapping with the auditory-evoked N1–P2 complex, for the microstates analyses we focused on the latency bands were N1 and P2 topographies occurred. In particular, this latency bands were defined as the time interval between the first occurrence of the N1 topography and the last manifestation of the P2 topography in the segmentation of the microstate analysis.

In order to evaluate any statistical differences in microstates duration and maximal GFP between groups, conditions, and runs, we labelled each time point of the single subject's ERPs according to the map which spatially best correlated with the template maps occurring in the time window of interest. It is important to mention again, that for both the clustering- and the refitting-processes, we adopted the temporal criterion of at least 10 time points; thus, a microstate had to persist for at least 20 ms. The statistical evaluation between the two groups was performed by using ANOVAs and based on the following independent variables: duration and maximal GFP of the microstates of interest.

Results

Behavioural Data

A generalized linear mixed model analysis for accuracy data and a mixed-effect regression model analysis for RT data were conducted by using *run* (run 1, 2, 3, 4), *condition* (ENV, FS), and *group* (musician, non-musician) as fixed factors, and *participants* as well as *items* as random factors. All post hoc *t* tests were corrected for multiple comparisons by using the Bonferroni procedure.

Accuracy

The generalized linear mixed model analysis for accuracy data revealed a significant effect of *condition* ($z = 8.604$; $P < 0.001$), which became manifest in terms of a higher accuracy for the identification of the FS stimuli, irrespective of group. Furthermore, we found a significant effect of *run* ($z = 4.802$; $P < 0.001$), which was associated with an increase in accuracy for both groups over the four runs, irrespective of condition. We also revealed a significant *condition by run* interaction effect ($z = 2.078$; $P = 0.038$. The *condition by run* interaction was manifested by a stronger increase in accuracy over the four runs in the FS condition compared to the ENV condition. Most interestingly, we found support for our working hypothesis in that the generalized linear mixed model analysis yielded a *condition by group by run* interaction ($z = 3.531$;

$P < 0.001$). In order to disentangle this three-way interaction, we performed two separate generalized linear mixed model analyses, one for each condition.

These further statistical analyses yielded a significant *group by run* interaction effect, but only in the FS condition (FS, $z = 4.099$; $P < 0.001$, ENV, $z = 1.320$; $P = 0.187$). Post hoc *t* tests (two-tailed) calculated for each run across the two groups in the FS condition, clearly showed that the behavioural performance differed between the two groups only during the second run; the musicians exhibited enhanced levels of accuracy as compared to the non-musicians (run 1, $T_{(48)} = -0.277$; $P = 0.783$; run 2, $T_{(48)} = -0.2.670$; $P = 0.010$; run 3, $T_{(48)} = 1.533$; $P = 0.127$; run 4, $T_{(48)} = 1.284$; $P = 0.205$).

In summary, the evaluation of accuracy revealed that (1) both groups showed a better behavioural performance during the FS condition; nevertheless, (2) the musicians' performance improved quicker than the non-musicians' performance during the FS condition. This means that the musicians' accuracy increased from the first to the second run faster than the controls' accuracy did.

Reaction Time

The mixed-effect model analysis for reaction time (RT) yielded significant effects of *condition* ($F_{(1,\ 5702)} = 296.51$; $P < 0.001$) and *run* ($F_{(3,\ 5603)} = 307.50$; $P < 0.001$). Furthermore, the same statistical approach revealed a significant *condition by group* interaction effect ($F_{(1,\ 5702)} = 18.289$; $P < 0.001$), as well as a *run by group* interaction ($F_{(3,\ 5603)} = 28.928$; $P < 0.001$). As visible in Fig. 1, both groups showed a perceptual learning effect during the course of the experiment that was characterized by a shorter reaction time. In line with the behavioural data resulting from the evaluation of accuracy, all participants experienced more difficulties when performing the ENV condition, which was associated with a longer reaction time.

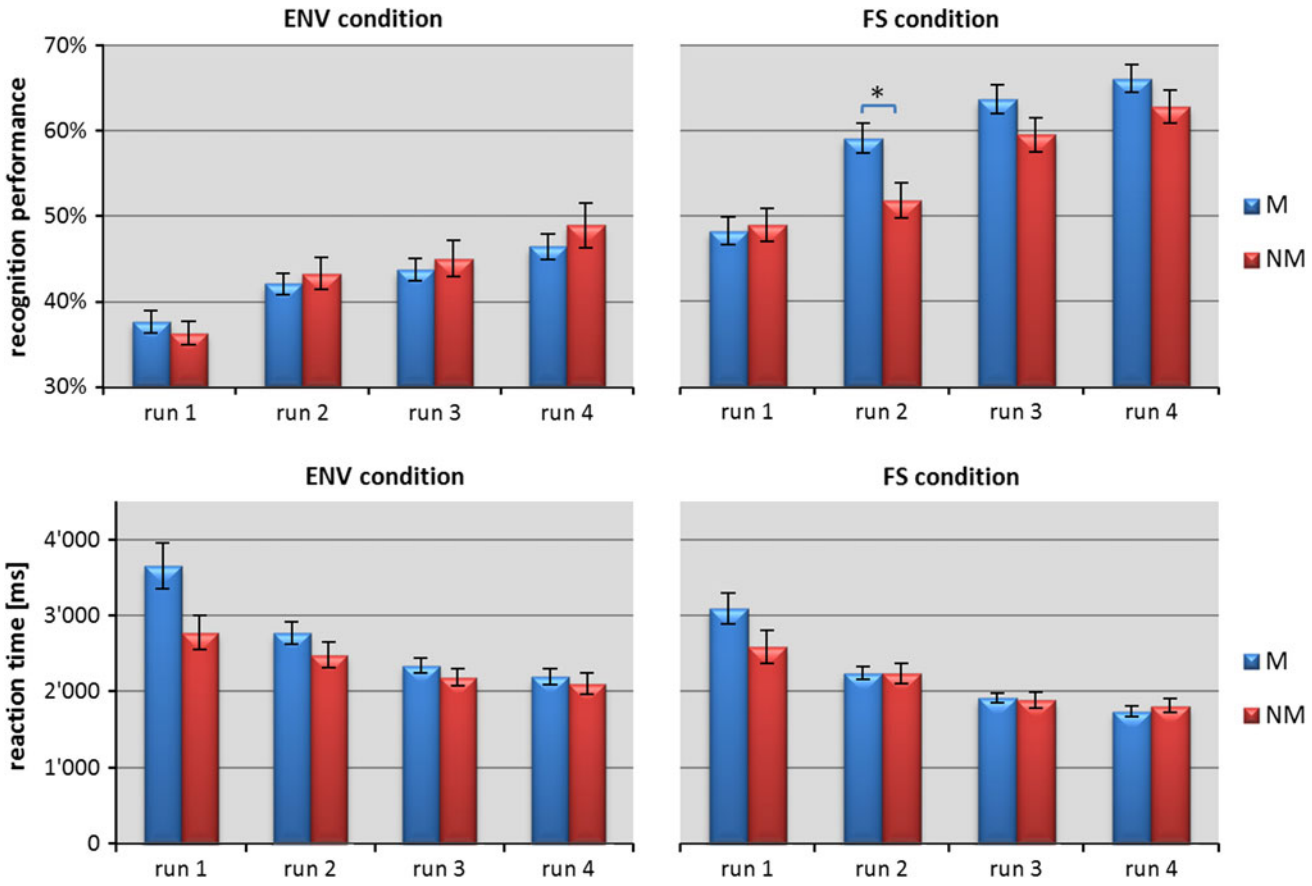

Fig. 1 Whereas the *top left panel* shows recognition performance in the ENV condition over the four runs for both groups, the *top right panel* depicts recognition performance in the FS condition. The *bottom left panel* displays the reaction time over all four runs and both groups during the ENV condition. The *bottom right panel* indicates the reaction time during the FS condition. Musicians are depicted in *blue*; non-musicians in *red*

EEG Data

Global Field Power (GFP)

In order to determine the time windows of interest for the microstate analyses, we computed GFP values for each participant by normalizing data level to the mean GFP for each subject. Group comparisons, which were evaluated by calculating *t* tests across the two groups, revealed a significant increase in GFP amplitude in the musician group for the latency band that ranged from 146 to 168 ms after stimulus onset (Fig. 2). This latency band overlaps with the temporal dynamic of the auditory evoked N1–P2 complex. Therefore, we performed the successive analysis (microstates) by focussing on the N1- and P2-like microstates.

Topographic Pattern Analyses: Microstates

The microstate analyses that we performed on the averaged AEPs of the two conditions (FS, ENV), the four runs (run 1, 2, 3, 4), and the two groups yielded seven representative topographic scalp voltage maps ("template maps"). Figure 3 displays these seven template maps with the associated GFP time courses. Due to the fact that we were specifically interested in the evaluation of the microstates overlapping with the time course of the N1–P2 complex, we restricted our analysis to N1- and P2-like topographic maps at around 100 and 200 ms. We identified "template map 4" as reflecting an N1-like microstate and "template map 5" as reflecting a P2-like microstate.

In the next processing step, we manually defined the exact time window for the microstate analyses based on the time period between the onset of the N1-like microstate and the offset of the P2-like microstate; namely, from 84 to 380 ms. Next, we refitted all template maps appearing within this time period back to the single subjects' data; that is, template maps 4, 5, 6, and 7. This specific procedure was performed, in order to obtain the duration and maximal GFP value of the topographic "template maps" associated with the N1 and P2 components for each subject, condition, and run. Statistical analyses were performed by computing four separate 4 × 2 × 2 ANOVAs for the following variables: duration, maximal GFP,

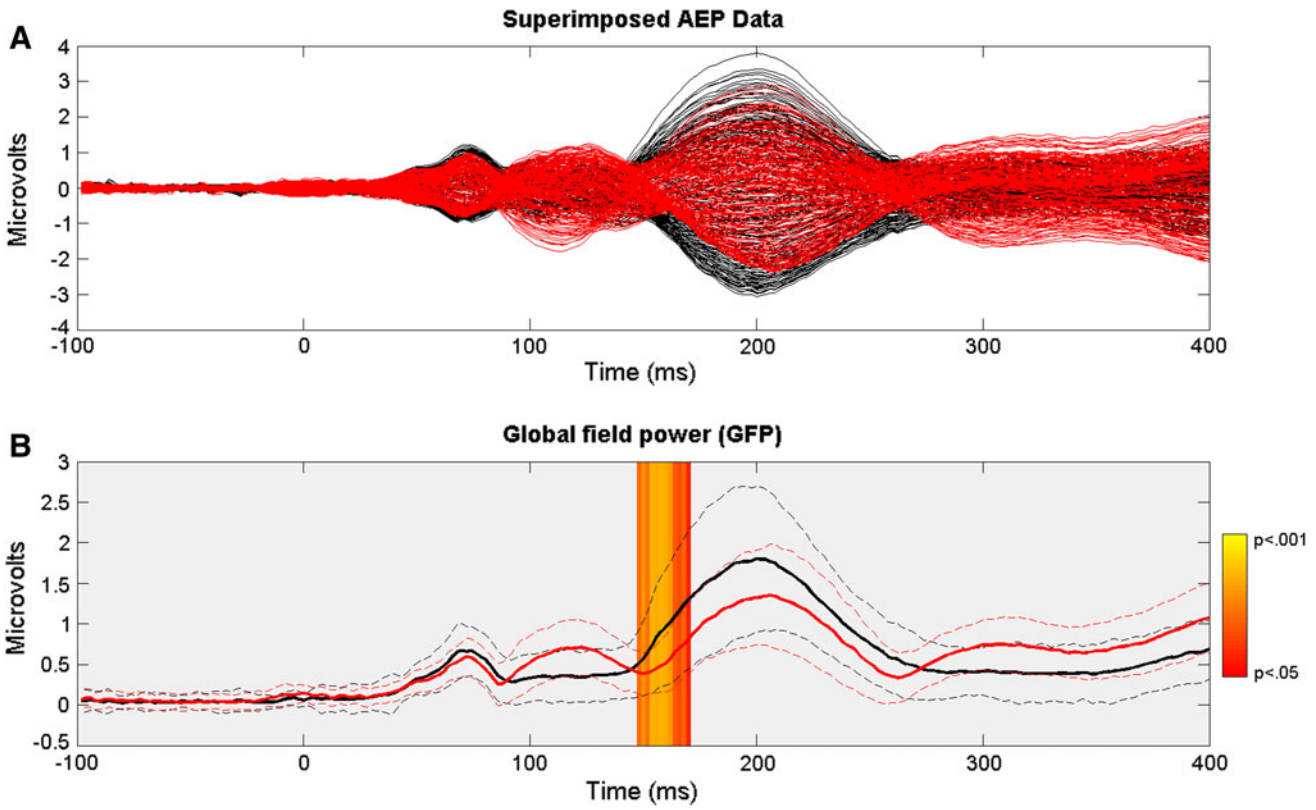

Fig. 2 **a** This figure shows separate superimposed AEPs for each of the 204 channels for musicians (*black*) and non-musicians (*red*). All conditions and runs were collapsed together. **b** Here GFP related to the musicians (*black*) and non-musicians (*red*) are displayed. The *dashed lines* indicate standard deviation. Significant GFP differences between the two groups are depicted in hot colours. The bar on the right side of image B depicts the significance threshold ranging from 0.05 to 0.001

N1-like microstates, and P2-like microstates. The factors *run* (run1, 2, 3, 4) and *condition* (ENV, FS) served as the within-subject factors, the factor *group* (musician, non-musician) functioned as the between-subject factor.

Duration of the N1-Like Microstate

Concerning the duration of the N1-like microstate, the 4 × 2 × 2 ANOVA (repeated measures) revealed a significant main effect of *group* ($F_{(1,48)} = 121.276$; $P < 0.001$; $\eta^2 = 0.158$), which was associated with a significant shorter duration of the N1-like microstate in the musician group. No further significant effects were found in this statistical analysis.

Duration of the P2-Like Microstate

For the duration of the P2-like microstate, the 4 × 2 × 2 ANOVA revealed a significant main effect of *group* ($F_{(1,48)} = 10.919$; $P = 0.002$; $\eta^2 = 0.185$) that was characterized by a significant longer duration of the P2-like microstate in the musicians, in comparison to the non-musicians. Furthermore, the same analysis revealed a main effect for *condition* ($F_{(1,48)} = 13.889$; $P < 0.001$; $\eta^2 = 0.224$). This main effect was associated with a longer duration of the P2-like microstate during the FS condition (Fig. 3). A slight trend was also observed for the factor *run* ($F_{(3,144)} = 3.088$; $P = 0.060$; $\eta^2 = 0.060$). All microstate results are depicted in Fig. 4.

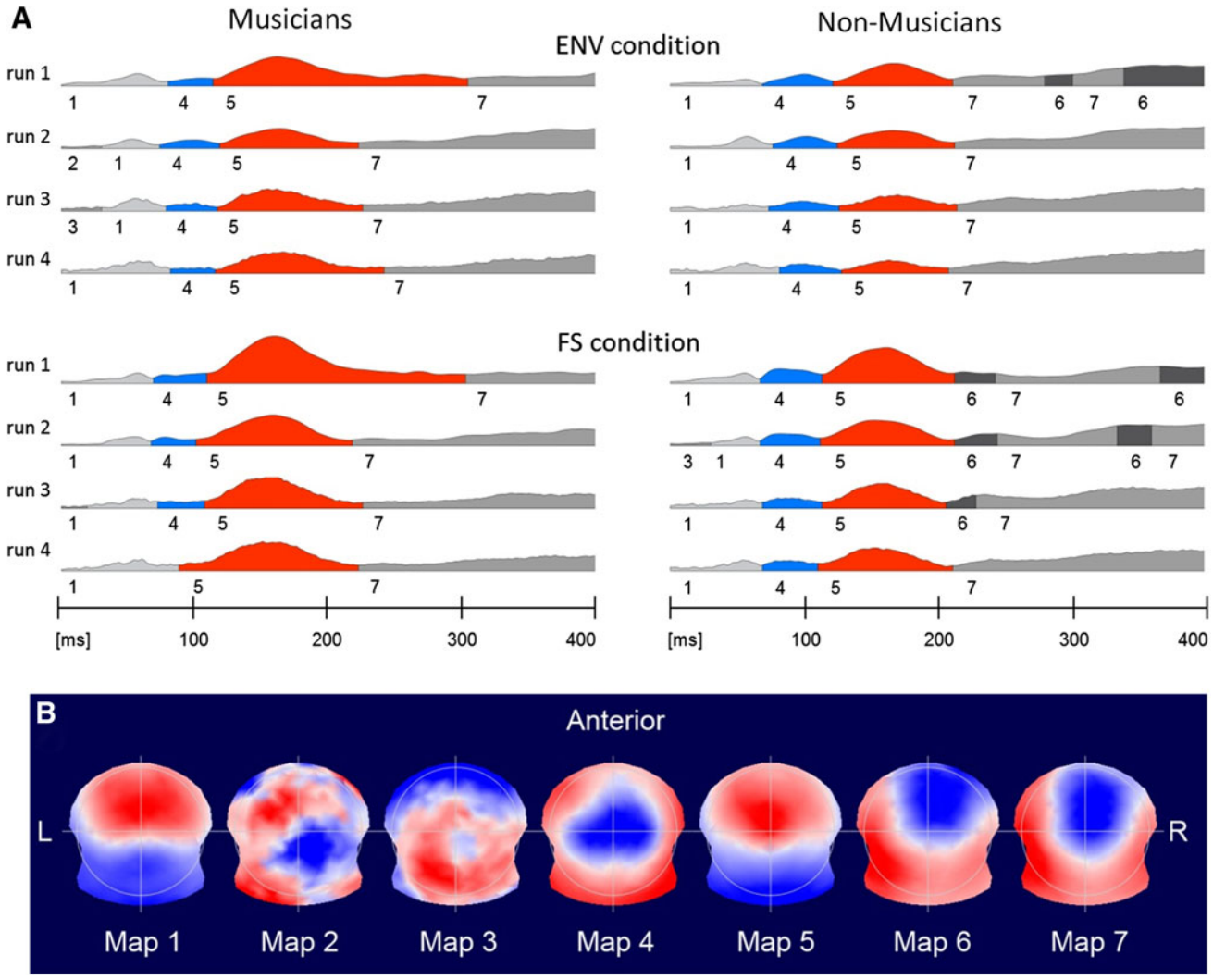

Fig. 3 This figure depicts the results of the microstate segmentation procedure. Seven different microstates were found to be most suitable for explaining the data. These seven topographic maps resulted from the microstate analyses that were calculated for the averaged AEPs in the two conditions (FS, ENV), four runs (run 1, 2, 3, 4) and two groups. Template Maps 4 and 5 represent the N1-like- and P2-like microstates, respectively

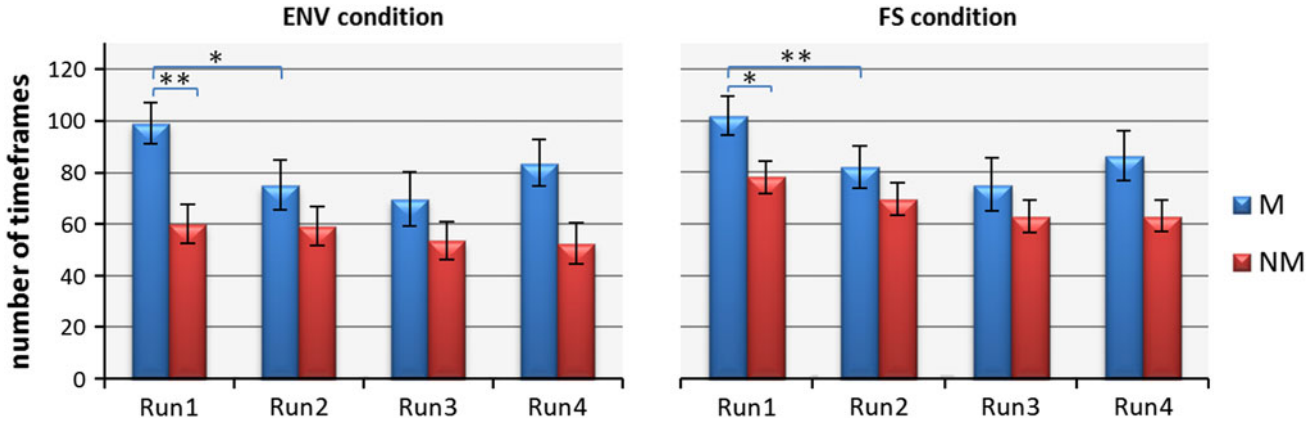

Fig. 4 This figure shows the duration of the P2-like microstate for each of the four runs and for each of the two groups. The *left panel* indicates the duration of the P2-like microstate during the ENV condition. The *right panel* illustrates the length of the P2-like microstate during the FS condition. Musicians are denoted by *blue*: non-musicians by *red*

By considering Fig. 4 it becomes apparent that the P2-like microstate duration fundamentally decreases from run 1 to run 2, but only in the musician group. During the next processing step, we compared the duration of the P2-like microstates across the first two runs and the two groups. In particular, we performed a 2 × 2 × 2 ANOVA analysis with *run* (run 1, 2) and *condition* (ENV, FS) as the within-subject factors and *group* (musicians, non-musicians) as the between-subject factor. This analysis yielded a significant main effect of *run* ($F_{(1,\ 48)} = 11.542$; $P < 0.001$; $\eta^2 = 0.194$), which was characterised by a longer P2-like microstate duration in the first run as compared to the second run, irrespective of group. The same statistical analysis also revealed a significant main effect of *group* ($F_{(1,\ 48)} = 6.494$; $P = 0.014$; $\eta^2 = 0.119$) that was associated with a longer P2-like microstate duration in the musician group in comparison to the non-musicians. We also found a significant *run by group* interaction effect ($F_{(1,\ 48)} = 4.815$; $P = 0.033$; $\eta^2 = 0.091$). Post hoc *t* tests revealed a significant longer duration of the P2-like microstates in run 1 than in run 2, but only within the musician group (musicians, $T_{(24)} = 3.710$; $P < 0.001$; non-musicians, $T_{(24)} = 0.915$; $P = 0.369$). In addition, musicians showed a significant longer duration of the P2-like microstate in run 1 in comparison to the non-musicians ($T_{(24)} = 3.292$; $P = 0.003$). By contrast, we did not reveal group differences in run 2 ($T_{(24)} = 1.370$; $P = 0.183$).

Maximal GFP Amplitude Related to the N1-Like Microstate

We computed a 4 × 2 × 2 ANOVA (4 runs, 2 conditions, and 2 groups; repeated measures), in order to evaluate the maximal GFP amplitude related to the N1-like microstate. This statistical approach yielded a significant main effect for *group* ($F_{(1,\ 48)} = 5.984$; $P = 0.018$; $\eta^2 = 0.111$) that was characterized by increased maximal GFP values in the non-musician group. All other main effects or interaction effects did not reach significance.

Maximal GFP Amplitude Related to the P2-Like Microstate

The evaluation of the maximal GFP amplitude related to the P2-like microstate, (4 × 2 × 2 ANOVA) our analysis revealed a significant main effect of *condition* ($F_{(1,\ 48)} = 125.933$; $P < 0.001$; $\eta^2 = 0.724$). The FS condition elicited increased maximal GFP values relative to the ENV condition (See Fig. 5). The same analysis resulted in a main effect of *run* ($F_{(3,\ 144)} = 6.948$; $P = 0.002$; $\eta^2 = 0.126$), which was characterized by a reduction in the maximal GFP during the time course of the experiment in both groups.

Additionally, we performed a less restricted statistical analysis, in order to evaluate whether perceptual learning effects affected the maximal GFP in the two groups during run 1 and run 2 differently. With this purpose in mind, we evaluated the maximal GFP values related to the P2-like microstate by performing a 2 × 2 × 2 ANOVA with the factors *run* (run 1, 2) and *condition* (ENV, FS) as the within-subject factors and *group* (musicians, non-musicians) as the between-subject factor. We found a significant main effect of *run* ($F_{(1,\ 48)} = 20.646$; $P < 0.001$; $\eta^2 = 0.301$), as well as a statistical trend for the *run by group* interaction ($F_{(1,\ 48)} = 2.981$; $P = 0.091$; $\eta^2 = 0.058$). Post hoc *t* tests revealed a significant decrease in the maximal GFP from run 1 to run 2 only in the musician group (musicians, $T_{(24)} = -4.223$; $P < 0.001$; non-musicians, $T_{(24)} = -2.013$; $P = 0.055$).

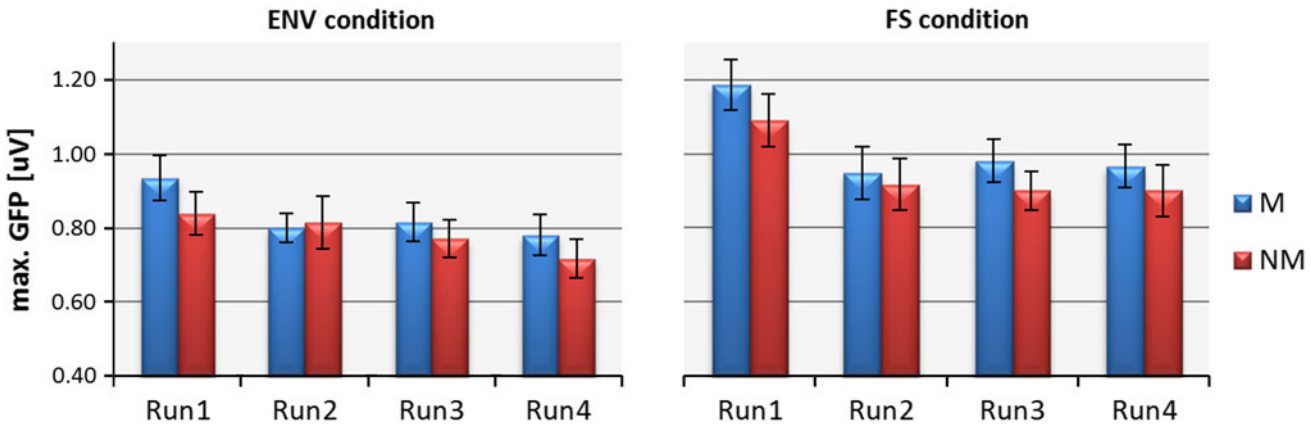

Fig. 5 This figure reveals the maximal global field power values (GFP) associated with the P2-like microstate for each of the four runs and the two groups. The *left panel* depicts the maximal GFP of the ENV condition. The *right panel* illustrates the maximal GFP of the FS condition. Musicians are depicted in *blue*; non-musicians in *red*

Discussion

General Discussion

Due to previous work showing that professional music training favours segmental (Marie et al. 2011; Besson et al. 2011; Elmer et al. 2012), as well as supra-segmental aspects of speech processing (Oechslin et al. 2010; Moreno et al. 2009), we designed an EEG study to specifically investigate the influence of musical expertise on perceptual verbal learning mechanisms. With this purpose in mind, we presented phonotactically legal pseudoword-chimeras to musically trained and untrained individuals. We created these stimuli by independently manipulating the fine structure and the envelope of the verbal stimuli. Based on converging results showing that professional music training improves the perception of spectral information in speech (Kraus et al. 2008; Kuriki et al. 2006; Meyer et al. 2006; Pantev et al. 2001b) and music sounds (Magne et al. 2006), we hypothesised that musicians will recognize an increased number of fine structure manipulated stimuli in the FS condition, compared to non-musicians. We also expected to find that long-term musical training facilitates the recognition and categorization of speech stimuli that are characterized by a reduced envelope structure. This can be expected even though there is only meagre evidence pointing in this direction. In terms of electrophysiological measurements, we expected to find evidence for microstate-differences. In particular, we hypothesized that there would be differences between the two groups in the fit between latency bands and the auditory evoked N1–P2 complex. We expected to observe these group-differences at least during the first half of the experimental session. In fact, it is logical to reason that musical experts, who are specifically trained to extract spectral- and envelope-related acoustic information from an auditory stream (for an overview consider Kraus et al. (2009)), will show faster perceptual learning effects than individuals who experience more difficulties when extracting similar information (non-musicians).

Behavioural Data

During the EEG experiment participants were instructed to assign the auditory stimuli, which were characterized by different acoustic manipulations, to one of four pseudowords that were visually presented on a monitor. This sort of task implies that the participants perceive the auditory stimuli and categorize them. The evaluation of the behavioural data showed that all participants increased their recognition performance and reaction time over the four runs of the experiment; this result clearly shows perceptual learning effects in both groups (Fig. 1). Our results revealed a main effect of condition (both accuracy and reaction time), which was irrespective of group and was manifested by an increased accuracy and shorter reaction time in the FS condition compared to the ENV condition. This evidences demonstrates that both groups experienced more difficulties in performing the ENV condition.

Furthermore, the evaluation of both reaction time and accuracy revealed significant interaction effects. Whereas the evaluation of RT yielded a significant condition by group interaction effect, the analysis of the accuracy data revealed a condition by run by group interaction. A further statistical evaluation of the three-way interaction effect (i.e., condition by run by group interaction) showed a significant run by group interaction effect, but only during the FS condition. Taken together, these results indicate that both groups generally showed a faster reaction time in response to the FS stimuli, and that the musicians' accuracy scores were significantly higher than those of the non-musicians during the FS condition. These results are in line

with previous work showing the profound influence of musical training on the processing and detection of spectral modulations in speech (Shannon et al. 1995; Smith et al. 2002) and non-speech signals (Baumann et al. 2008; Kuriki et al. 2006; Meyer et al. 2011; Pantev et al. 2001a; Shahin et al. 2005). Our results are novel in that we provide evidence for the notion that musical training facilitates the perception and therefore the recognition of pseudowords, which were manipulated in terms of spectral information and not included envelope-related information. This result may support the view that musicians rely on spectral information more strongly than non-musicians for decoding speech signals. This may be associated with the fact that it is principally the spectral-, and not the envelope-related information, which forms the basic foundation of music practice.

This study has produced evidence for the existence of perceptual verbal learning effects in the musician group during the FS condition. This effect was manifested by an increase in accuracy from the first to the second run of the experiment. This suggests that musical training favours the perceptual learning of pseudoword-chimeras characterized by spectral information. This specific result supplements the previously described advantage of musicians in processing spectral speech information. In fact, we not only provide evidence for the fact that musicians are more accurate in recognizing spectral information in speech signals, but also that they learn faster than non-musicians. This result is relevant as it provides evidence for transfer effects from musical training to the perception of spectral information in speech signals (Schneider et al. 2005b).

Electrophysiological Data

The Effect of Experimental Conditions

The electrophysiological data are in line with the aforementioned behavioural results. We revealed a main effect of condition, but only in latency bands overlapping with the P2-related component. In both groups, the FS condition elicited stronger maximal GFP amplitudes and was associated with a longer duration of the P2-like microstate, in comparison to the stimuli pertaining to the ENV condition. From a psychoacoustic perspective it is obvious that the fundamental difference between the FS and ENV conditions is that the ENV condition does not contain any spectral information. In this context, previous electrophysiological work has shown that spectrally more complex acoustic sounds are generally associated with an increase in the N1–P2-related amplitudes (Meyer et al. 2006). An earlier EEG study by Shahin et al. (2005) evidenced that it was not the N1 amplitude, which reflected the spectral complexity when the temporal envelope was held constant, but rather the P2 amplitude. This previous work clearly points to different spectral encoding mechanisms underlying N1 and P2 responses.

Even though our results are in line with previous work showing a particular sensitivity of P2-related latency bands to the processing of spectral information (Baumann et al. 2008; Kraus et al. 2009; Magne et al. 2006; Meyer et al. 2006; Shahin et al. 2003; Trainor et al. 2003), we provide additional information regarding the topographical patterns associated with spectral processing. In fact, by applying the topographic pattern analysis proposed by Murray et al. (2008), we provide evidence that spectral complexity profoundly influences the maximal GFP amplitude, especially as regards the duration of the P2-like microstate. This result suggests that the duration of the P2-like microstate can be used as an alternative electrophysiological marker for spectral complexity. Thus, a promising approach in the future will be to reassess whether this specific electrophysiological marker has the potential to be used for differentiating between normally hearing people and individuals with specific hearing impairments. In fact, it has been proposed that microstates, which vary in topographical, putative ‘‘atoms of thought’’, may represent different types of information processing (Schlegel et al. 2011).

The General Effect of Musical Expertise

The evaluation of the N1-like microstate yielded a significant shorter duration, as well as a decreased maximal GFP, in the musician group as compared to the non-musicians. In addition, the evaluation of the P2-like microstate revealed a longer P2-like microstate duration in the musician group only. All these electrophysiological results were not influenced by run or condition.

These results are consistent with previous studies showing differential electrical brain responses in latency bands corresponding to the auditory N1 and P2 components between musically skilled and unskilled individuals during speech- (Marie et al. 2011; Ott et al. 2011) and non-speech (Pantev et al. 1998) processing. Certainly, our study cannot be directly compared with those mentioned above because instead of presenting musical items or natural speech sounds to the participants, we presented pseudoword-chimeras. A further novel aspect of our study is that we are the first research group to focus on microstate analyses for investigating speech processing and perceptual learning mechanisms in professional musicians and non-musicians (Ott et al. 2011).

Meanwhile, it is established that N1- and P2-like topographic maps, which are typically characterized by polarity reversal around the mastoid regions, reflect electrical

activity originating from auditory-related brain regions (Vaughan and Ritter 1970). A similar tuning of auditory-related brain regions in response to speech stimuli as a function of expertise was described by other groups (Marie et al. 2011; Ott et al. 2011). Since these differential brain responses between the two groups were irrespective of condition and run, microstates results are interpreted as reflecting an altered sensitivity of musicians for processing acoustic information in general (Kraus and Chandrasekaran 2010), and speech sounds in particular (Besson et al. 2011; Meyer et al. 2007; Shahin 2011).

To date there is only sparse literature describing the influence of perceptual and cognitive processes on the spatiotemporal dynamics of the N1–P2 complex (Besson et al. 2011). Therefore, we can only speculate about the intrinsic meaning of longer or shorter microstate durations in conjunction with musical practice and expertise. Otherwise, there is some evidence indicating that top-down processes are often associated with the modulation of the auditory-evoked P2 component (Picton and Hillyard 1974; Hillyard 1981; Schwent and Hillyard 1975). Furthermore, it is suggested that the N1 component is more likely influenced by the physical attributes of the acoustic signals than by other cognitive parameters. Thus, it is possible that our results reflect a differential engagement of perceptual and cognitive strategies between the two groups as a function of musical expertise. In particular, it is plausibly to think that the intensively trained auditory system of musicians enables a faster and more efficient perceptual analysis of the verbal stimuli. This is reflected by a shorter N1-related microstate duration and a reduced maximal GFP amplitude. This specific pattern of brain response may be particularly advantageous for a faster allocation of cognitive resources, which are necessary for performing the task. Although further studies are necessary for better understanding the relationship between microstate duration, perception, and cognition; we consider the topographical pattern analysis presented by Murray and colleagues (2008) as a powerful and novel approach for investigating the electrophysiological correlates of musicianship.

Musical Expertise Bolsters Perceptual Learning Mechanisms

The electrophysiological evidence we found for faster learning effects in the musician group is in line with the aforementioned behavioural data. This perceptual learning effect was reflected in the FS condition by a shortening of the P2-like microstate duration. This is of particular interest because it reveals that perceptual learning effects are reflected by the duration of the microstate associated with the P2-component. We found that microstate duration can be used as a marker for acoustic expertise, as well as for evaluating perceptual learning effects. We can only speculate whether this superior perceptual performance is accompanied by a more efficient engagement of top-down resources, which are necessary for accomplishing this task. Even though we are fully aware that further research is necessary for better understanding which perceptual and cognitive processes are exactly reflected by the microstates duration, our electrophysiological data are in line with previous work indicating that perceptual processing is principally reflected by a modulation of the electrophysiological responses overlapping with the event-related P2 processing stage (Alain et al. 2010; Ben-David et al. 2010).

Our experiment revealed a differential modulation of microstates between the two groups in latency bands overlapping with the time course of the P2 component and is assumed to be associated with increased perceptual skills as a function of musicianship. In fact, the musician's increased behavioural performance during the first two runs was accompanied by a reduction of the P2-like microstate duration, as well as a reduction of its maximal GFP amplitude. We observed a decrease of the P2-like microstate duration during the entire experiment in both groups, which was associated with an increased behavioural performance. This finding supports the hypothesis that the duration of the P2-like microstate can be used as a marker for perceptual learning mechanisms.

Finally, it should be mentioned that in the present study we presented disyllabic pseudowords chimeras to a group of musicians and non-musicians in order to investigate the influence of auditory expertise on perceptual verbal learning mechanisms. Basing on previous evidence showing that the increased auditory acuity of professional musicians is primarily reflected by brain responses overlapping with the time course of the N1–P2 complex (Ott et al. 2011; Marie et al. 2011), in the present work we only focused on these latency bands. From a pragmatic point of view, this means that our results are restricted to the processing of the first syllable of the pseudowords. Otherwise, one should consider that the disyllabic words we used were composed of two elements with the same physical modulation, which is spectral or temporal. This implies that the analysis of the first syllable is representative for the processing of perceptual stimulus attributes. Furthermore, in the present work we consciously decided to present phonotactically legal disyllabic pseudowords instead of syllables because the former are indeed more likely similar to real speech. Since the present work primarily focused on perceptual and not cognitive aspects of expertise, we were not interested in evaluating late processing stages. Certainly, further studies are necessary for better elucidate the relationship between microstate duration and different learning mechanisms.

Limitations

The present study was designed to investigate the effect of musical expertise on the perceptual learning of acoustically-reduced pseudoword-chimeras. Even though this study provides novel insights, a few limitations should be noted. Although we provide evidence that microstates can be used as a marker for musical expertise and learning effects, it is nevertheless difficult to explain which perceptual and cognitive processes are exactly reflected by this measure. Further studies will be useful to shield light on the influence of bottom-up and top-down processes on microstates duration, as well as to distinguish between different learning processes. A further limitation of our work is that the sample of musicians we measured was quite heterogeneous. We cannot exclude the possibility that the singers in our sample, who constituted a quarter of the entire group, may have influenced the data in some direction. Further studies are necessary to determine whether the primary instrument played by a musician may have an influence on transfer effects from musical training to speech processing. Finally, even though this study only evaluated perceptual processes restricted to latency bands overlapping with the N1–P2 complex which was previously shown to constitute an objective marker of auditory processing, further studies are necessary in order to better understand the microstates associated with subsequent stages of speech processing.

Acknowledgments We thank Sarah McCourt Meyer for comments on a previous version of the manuscript. This work was supported by Swiss National Foundation (320030-120661 and 4-62341-08).

Conflict of interest None.

References

Alain C, Campeanu S, Tremblay K (2010) Changes in sensory evoked responses coincide with rapid improvement in speech identification performance. J Cogn Neurosci 22(2):392–403. doi:10.1162/jocn.2009.21279

Aleman A, Nieuwenstein MR, Bocker KB, de Haan EH (2000) Music training and mental imagery ability. Neuropsychologia 38(12):1664–1668

Annett J (1970) Role of action feedback in acquisition of simple motor responses. J Motor Behav 2(3):217–221

Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. J Mem Lang 59(4):390–412. doi:10.1016/j.jml.2007.12.005

Baumann S, Meyer M, Jancke L (2008) Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. J Cogn Neurosci 20(12):2238–2249. doi:10.1162/jocn.2008.20157

Ben-David BM, Campeanu S, Tremblay KL, Alain C (2010) Auditory evoked potentials dissociate rapid perceptual learning from task repetition without learning. Psychophysiology. doi:10.1111/j.1469-8986.2010.01139.x

Besson M, Schon D, Moreno S, Santos A, Magne C (2007) Influence of musical expertise and musical training on pitch processing in music and language. Restor Neurol Neurosci 25(3–4):399–410

Besson M, Chobert J, Marie Cl (2011) Transfer of training between music and speech: common processing, attention and memory. Frontiers in Psychology 2: 94. doi:10.3389/fpsyg.2011.00094

Bilhartz TD, Bruhn RA, Olson JE (1999) The effect of early music training on child cognitive development. J Appl Dev Psychol 20(4):615–636

Boh B, Herholz SC, Lappe C, Pantev C (2011) Processing of complex auditory patterns in musicians and nonmusicians. PLoS One 6(7):e21458. doi:10.1371/journal.pone.0021458.PONE-D-11-01681

Brochard R, Dufour A, Despres O (2004) Effect of musical expertise on visuospatial abilities: evidence from reaction times and mental imagery. Brain Cogn 54(2):103–109. doi:10.1016/S0278-2626(03)00264

Brysbaert M (2007) "The language-as-fixed-effect fallacy": Some simple SPSS solutions to a complex problem. University of London, Royal Holloway

Clark HH (1973) Language as fixed-effect fallacy: critique of language statistics in psychological research. J Verb Learn Verb Be 12(4):335–359

Draganova R, Wollbrink A, Schulz M, Okamoto H, Pantev C (2009) Modulation of auditory evoked responses to spectral and temporal changes by behavioral discrimination training. BMC Neurosci 10:143. doi:10.1186/1471-2202-10-143

Elmer S, Meyer M, Jancke L (2012) Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. Cereb Cortex 22(3):650–658. doi:bhr142 [pii] 10.1093/cercor/bhr142

Fort A, Delpuech C, Pernier J, Giard MH (2002) Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. Cereb Cortex 12(10):1031–1039

Gaab N, Tallal P, Kim H, Lakshminarayanan K, Archie JJ, Glover GH, Gabrieli JD (2005) Neural correlates of rapid spectrotemporal processing in musicians and nonmusicians. Ann N Y Acad Sci 1060:82–88. doi:10.1196/annals.1360.040

Gordon E (1989) Advanced measures of music audition. GIA Publications, Chicago

Hillyard SA (1981) Selective auditory attention and early event-related potentials: a rejoinder. Can J Psychol 35(2):159–174

Ho Y-C, Cheung M-C, Chan AS (2003) Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. Neuropsychology 17(3):439–450

Hyde M (1997) The N1 response and its applications. Audiol Neurootol 2(5):281–307

Hyde KL, Lerch J, Norton A, Forgeard M, Winner E, Evans AC, Schlaug G (2009) The effects of musical training on structural brain development: a longitudinal study. Ann N Y Acad Sci 1169:182–186. doi:10.1111/j.1749-6632.2009.04852.x

Jancke L (1996) The hand performance test with a modified time limit instruction enables the examination of hand performance asymmetries in adults. Percept Motor Skill 82(3):735–738

Jancke L (2009) The plastic human brain. Restor Neurol Neurosci 27(5):521–538. doi:10.3233/RNN-2009-0519

Jung TP, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui V, Sejnowski TJ (2000) Removing electroencephalographic artifacts by blind source separation. Psychophysiology 37(2):163–178

Kraus N, Chandrasekaran B (2010) Music training for the development of auditory skills. Nat Rev Neurosci 11(8):599–605. doi:10.1038/nrn2882

Kraus N, Skoe E, Parbery-Clark A (2008) Auditory processing of pitch, timbre and time: implications for language and music. Paper presented at the 2008 Research symposium: hear our voices, Milwaukee

Kraus N, Skoe E, Parbery-Clark A, Ashley R (2009) Experience-induced malleability in neural encoding of pitch, timbre, and timing. Ann N Y Acad Sci 1169:543–557. doi:10.1111/j.1749-6632.2009.04549.x

Krzanowski WJ, Lai YT (1988) A criterion for determining the number of groups in a data set using sum-of-squares clustering. Biometrics 44(1):23–34

Kuriki S, Kanda S, Hirata Y (2006) Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. J Neurosci 26(15):4046–4053. doi:10.1523/JNEUROSCI.3907-05.2006

Lee KM, Skoe E, Kraus N, Ashley R (2009) Selective subcortical enhancement of musical intervals in musicians. J Neurosci 29(18):5832–5840. doi:10.1523/JNEUROSCI.6133-08.2009

Lehmann D, Michel CM (2011) EEG-defined functional microstates as basic building blocks of mental processes. Clin Neurophysiol 122(6):1073–1074. doi:10.1016/j.clinph.2010.11.003

Lehrl S (1977) Mehrfachwahl-wortschatz intelligenz test (MWT-B). Perimed, Erlangen

Lehrl S, Gallwitz A, Blaha V, Fischer B (1991) Theorie und Messung der geistigen Leistungsfähigkeit mit dem Kurztest KAI. Vless, Ebersberg

Locker L, Hoffman L, Bovaird JA (2007) On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. Behav Res Methods 39(4):723–730

Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54(6):1001–1010. doi:10.1016/j.neuron.2007.06.004

Magne C, Schon D, Besson M (2006) Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. J Cogn Neurosci 18(2):199–211. doi:10.1162/089892906775783660

Marie C, Magne C, Besson M (2011) Musicians and the metric structure of words. J Cogn Neurosci 23(2):294–305. doi:10.1162/jocn.2010.21413

Marques C, Moreno S, Castro SL, Besson M (2007) Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. J Cognit Neurosci 19(9):1453–1463

Meyer M, Baumann S, Jancke L (2006) Electrical brain imaging reveals spatio-temporal dynamics of timbre perception in humans. NeuroImage 32(4):1510–1523. doi:10.1016/j.neuroimage.2006.04.193

Meyer M, Elmer S, Baumann S, Jancke L (2007) Short-term plasticity in the auditory system: differential neural responses to perception and imagery of speech and music. Restor Neurol Neurosci 25(3–4):411–431

Meyer M, Elmer S, Ringli M, Oechslin MS, Baumann S, Jancke L (2011) Long-term exposure to music enhances the sensitivity of the auditory system in children. Eur J Neurosci 34(5):755–765. doi:10.1111/j.1460-9568.2011.07795.x

Michel CM, Koenig Th, Brandeis D, Gianotti LRR, Wackermann J (2009) Electrical neuroimaging in the time domain: microstate analysis. In: Michel CM, Koenig Th, Brandeis D, Gianotti LRR, Wackermann J (eds) Electrical neuroimaging. Cambridge University Press, Cambridge, pp 123–125

Milovanov R, Tervaniemi M (2011) The interplay between musical and linguistic aptitudes: a review. Frontiers in Psychol 2. doi:10.3389/fpsyg.2011.00321

Moreno S, Marques C, Santos A, Santos M, Castro SL, Besson M (2009) Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. Cereb Cortex 19(3):712–723. doi:10.1093/cercor/bhn120

Munte TF, Altenmuller E, Jancke L (2002) The musician's brain as a model of neuroplasticity. Nat Rev Neurosci 3(6):473–478

Murray MM, Wylie GR, Higgins BA, Javitt DC, Schroeder CE, Foxe JJ (2002) The spatiotemporal dynamics of illusory contour processing: combined high-density electrical mapping, source analysis, and functional magnetic resonance imaging. J Neurosci 22(12):5055–5073

Murray MM, Michel CM, Grave de Peralta R, Ortigue S, Brunet D, Gonzalez Andino S, Schnider A (2004) Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. NeuroImage 21(1):125–135

Murray MM, Brunet D, Michel CM (2008) Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr 20(4):249–264. doi:10.1007/s10548-008-0054-5

Oechslin MS, Meyer M, Jancke L (2010) Absolute pitch–functional evidence of speech-relevant auditory acuity. Cereb Cortex 20(2):447–455. doi:10.1093/cercor/bhp113

Ostroff JM, Martin BA, Boothroyd A (1998) Cortical evoked response to acoustic change within a syllable. Ear Hearing 19(4):290–297

Ott CG, Langer N, Oechslin M, Meyer M, Jancke L (2011) Processing of voiced and unvoiced acoustic stimuli in musicians. Front Psychol 2:195. doi:10.3389/fpsyg.2011.00195

Pantev C, Oostenveld R, Engelien A, Ross B, Roberts LE, Hoke M (1998) Increased auditory cortical representation in musicians. Nature 392(6678):811–814. doi:10.1038/33918

Pantev C, Engelien A, Candia V, Elbert T (2001a) Representational cortex in musicians. Plastic alterations in response to musical practice. Ann N Y Acad Sci 930:300–314

Pantev C, Roberts LE, Schulz M, Engelien A, Ross B (2001b) Timbre-specific enhancement of auditory cortical representations in musicians. NeuroReport 12(1):169–174

Patel AD (2011) Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. Frontiers in Psychology 2. doi:10.3389/fpsyg.2011.00142

Perrin F, Pernier J, Bertrand O, Giard MH, Echallier JF (1987) Mapping of scalp potentials by surface spline interpolation. Electroen Clin Neuro 66(1):75–81

Picton TW, Hillyard SA (1974) Human auditory evoked potentials. II. Effects of attention. Electroencephalogr Clin Neurophysiol 36(2):191–199

Pratt H, Starr A, Michalewski HJ, Bleich N, Mittelman N (2007) The N1 complex to gaps in noise: effects of preceding noise duration and intensity. Clin Neurophysiol 118(5):1078–1087. doi:10.1016/j.clinph.2007.01.005

Rajan R, Irvine DR, Wise LZ, Heil P (1993) Effect of unilateral partial cochlear lesions in adult cats on the representation of lesioned and unlesioned cochleas in primary auditory cortex. J Comp Neurol 338(1):17–49. doi:10.1002/cne.903380104

Recanzone GH, Schreiner CE, Merzenich MM (1993) Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. J Neurosci 13(1):87–103

Schlaug G (2001) The brain of musicians. A model for functional and structural adaptation. Ann N Y Acad Sci 930:281–299

Schlaug G, Jancke L, Huang Y, Steinmetz H (1995) In vivo evidence of structural brain asymmetry in musicians. Science 267(5198):699–701

Schlaug G, Norton A, Overy K, Winner E (2005) Effects of music training on the child's brain and cognitive development. Ann N Y Acad Sci 1060:219–230. doi:10.1196/annals.1360.015

Schlegel F, Lehmann D, Faber PL, Milz P, Gianotti LR (2011) EEG microstates during resting represent personality differences. Brain Topogr. doi:10.1007/s10548-011-0189-7

Schneider P, Sluming V, Roberts N, Bleeck S, Rupp A (2005a) Structural, functional, and perceptual differences in Heschl's gyrus and musical instrument preference. Ann N Y Acad Sci 1060:387–394. doi:10.1196/annals.1360.033

Schneider P, Sluming V, Roberts N, Scherg M, Goebel R, Specht HJ, Dosch HG, Bleeck S, Stippich C, Rupp A (2005b) Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. Nat Neurosci 8(9):1241–1247. doi:10.1038/nn1530

Schön D, Francois C (2011) Musical expertise and statistical learning of musical and linguistic structures. Frontiers in Psychology 2. doi:10.3389/fpsyg.2011.00167

Schon D, Magne C, Besson M (2004) The music of speech: music training facilitates pitch processing in both music and language. Psychophysiology 41(3):341–349. doi:10.1111/1469-8986.00172.X

Schwent VL, Hillyard SA (1975) Evoked potential correlates of selective attention with multi-channel auditory inputs. Electroencephalogr Clin Neurophysiol 38(2):131–138

Shahin AJ (2011) Neurophysiological influence of musical training on speech perception. Front Psychol 2:126. doi:10.3389/fpsyg.2011.00126

Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE (2003) Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. J Neurosci 23(13):5545–5552

Shahin A, Roberts LE, Pantev C, Trainor LJ, Ross B (2005) Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. NeuroReport 16(16):1781–1785. doi:00001756-200511070-00011

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270(5234):303–304

Sharma A, Marsh CM, Dorman MF (2000) Relationship between N1 evoked potential morphology and the perception of voicing. J Acoust Soc Am 108(6):3030–3035

Shen C, Smith ZM, Oxenham AJ, Delgutte B (2001) Auditory chimera demo. http://epl.meei.harvard.edu/~bard/chimeral

Sluming V, Brooks J, Howard M, Downes JJ, Roberts N (2007) Broca's area supports enhanced visuospatial cognition in orchestral musicians. J Neurosci 27(14):3799–3806. doi:10.1523/Jneurosci.0147-07.2007

Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. Nature 416(6876):87–90. doi:10.1038/416087a

Sturm W, Willmes K (1999) Verbaler Lerntest und Nonverbaler Lerntest (VLT/NVLT). Hogrefe, Göttingen

Tibshirani R, Walther G (2005) Cluster validation by prediction strength. J Comput Graph Stat 14(3):511–528. doi:10.1198/106186005x59243

Trainor LJ, Shahin A, Roberts LE (2003) Effects of musical training on the auditory cortex in children. Ann N Y Acad Sci 999:506–513

Vaughan HG Jr, Ritter W (1970) The sources of auditory evoked responses recorded from the human scalp. Electroencephalogr Clin Neurophysiol 28(4):360–367